

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/41133>

Please be advised that this information was generated on 2018-07-07 and may be subject to change.

De Recognition Based Vocabulary Test: een nieuwe maat voor receptieve woordenschatkennis?

J. EYCKMANS (ERASMUS HOGESCHOOL BRUSSEL), H. VAN DE VELDE
(UNIVERSITEIT UTRECHT), R. VAN HOUT (UNIVERSITEIT NIJMEGEN)
& FR. BOERS (ERASMUS HOGESCHOOL BRUSSEL)

Une partie considérable de la recherche récente en matière d'acquisition de vocabulaire s'est consacrée à mesurer l'empan du vocabulaire réceptif en L2. Un outil fort sollicité dans ce cadre est le "Yes/No Vocabulary Test" (Meara & Buxton 1987). Cette épreuve invite l'apprenant en langue étrangère à indiquer, dans une liste de mots et de pseudo-mots, les mots dont il connaît la signification.

Le présent article expose pourquoi cette épreuve ne produit pas de résultats fiables dans le cas d'apprenants francophones du néerlandais (Eyckmans 2000, Beeckmans et al 2001). Les données expérimentales récoltées révèlent un biais dans les réponses des sujets, ce qui met en cause la validité du format. Les causes de ce biais sont analysées et un nouveau format d'évaluation est proposé, visant à contourner le "response bias". Dans cette nouvelle épreuve l'apprenant doit accomplir la tâche d'indiquer le mot existant dans une paire comprenant un mot et un pseudo-mot. Nous présentons les résultats d'une expérience dans laquelle 177 apprenants francophones du néerlandais ont effectués les deux tests. Leurs réponses furent validées à l'aide d'une épreuve de traduction. En conclusion, nous préconisons des formats d'évaluation lexicaux robustes par rapport aux caractéristiques de différentes populations.

1. Inleiding

In de boeken over woordenschatverwerving die de jongste jaren verschenen, wordt nadrukkelijk gesteld dat kennis van de meest frequente woorden van de doeltaal een noodzakelijke voorwaarde vormt voor het succes van het verdere taalleerproces (Nation 1990, 2001; Meara 1996; Schmitt 2000). Het verband tussen receptieve woordenschatomvang en de verschillende vaardigheden in een vreemde taal werd immers meermaals aangetoond in onderzoeksexperimenten (Hirsch & Nation 1992; Laufer 1992; Coady, Magoto, Hubbard, Graney & Mokhtari 1993). Tot op vandaag hebben twee toetsformaten opgang gemaakt als gestandaardiseerde toets voor receptieve woordenschatkennis: de *Vocabulary Levels Test* (Nation 1983, 1990) en de *Yes/No*

Vocabulary Test (Meara & Buxton 1987). In deze bijdrage zullen we het hebben over de problemen die opduiken wanneer de *Yes/No Vocabulary Test* wordt gebruikt voor Franstalige leerders van het Nederlands en stellen we een nieuw toetsformaat voor, de *Recognition Based Vocabulary Test* (Eyckmans, 2004). Na een beschrijving van de Ja/Nee Toets (Sectie 2), wordt de nieuwe woordenschattoets toegelicht (Sectie 3) en brengen we verslag uit van een experiment waarin beide toetsen werden afgelegd door dezelfde participanten (Sectie 4). Tot slot evalueren we de nieuwe toets op basis van de vergaarde empirische gegevens en maken we enkele kanttekeningen bij het gebruik van receptieve woordenschattoetsen (Sectie 5).

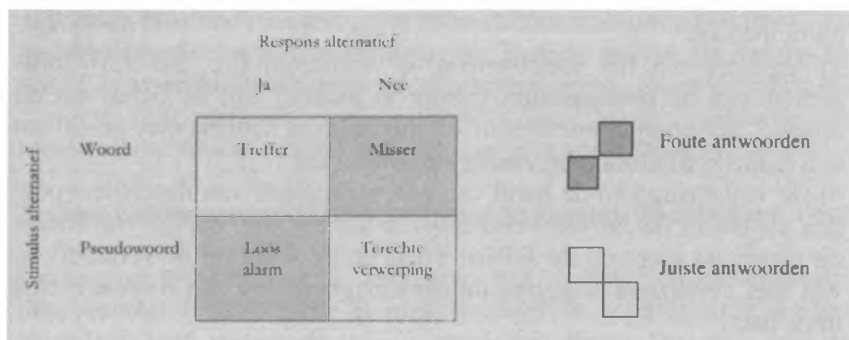
2. De Ja/Nee Woordenschattoets

2.1. Het toetsformaat

In de Ja/Nee woordenschattoets wordt de taalleerder een lijst van woorden uit de doeltaal voorgelegd die werd aangevuld met een aantal pseudowoorden (in een verhouding van 60 woorden en 40 pseudowoorden volgens de oorspronkelijke toets van Meara & Buxton 1987). De leerder wordt gevraagd met *Ja* of *Nee* aan te geven of hij de betekenis van deze woorden kent. In de meest recente versies van de toets werd deze opgave omgevormd tot de vraag of de woorden uit de lijst bestaan in de doeltaal of niet (Dialang 2003). De woorden en pseudowoorden worden geselecteerd binnen een bepaalde frequentieband en op basis van extrapolatie van de toetsscore wordt bepaald in welke mate de leerder de woordenschat uit die frequentieband receptief kent. Door verschillende toetsen (opgemaakt aan de hand van verschillende frequentiebanden) te laten maken, kan er een profiel worden opge maakt van de receptieve woordenschatkennis van een individu in om het even welke taal (Meara 1990, 1992). De DIALANG toetsbatterij (<http://www.dialang.org>) die werd aangelegd op initiatief van The Council of Europe om individuen in staat te stellen via een website hun taalvaardigheden in 14 Europese talen in te schatten, selecteerde de Ja/Nee Toets als gestandaardiseerde maat om de woordenschatkennis van de participanten in kaart te brengen.

De scoreberekening van de Ja/Nee Toets is geen eenvoudige procedure. Er zijn immers vier soorten stimulus-respons combinaties mogelijk aangezien de leerder wordt blootgesteld aan twee soorten stimuli (woorden en pseudowoorden) en telkens de keuze heeft tussen een *Ja* of een *Nee* antwoord (zie Figuur 1). In de literatuur werden verschillende correctieformules op de ruwe scores voorgesteld (Huibregtse, Admiraal & Meara 2001, Beeckmans, Eyckmans, Janssens, Dufranne & Van de Velde 2001), sommige ervan gebaseerd op Si-

gnaal Detectie Modellen, maar er bestaat voorlopig geen consensus over welke van deze formules het meest geschikt is. Het zou ons te verleiden in te gaan op deze discussie, het volstaat op te merken dat alle voorgestelde formules ervan uitgaan dat de *Treffer*-score moet aangepast worden aan de hand van de *Loos alarm*-score. De premisse waarvan wordt uitgegaan is dat wanneer twee participanten eenzelfde *Treffer*-score behalen, diegene met een hogere *Loos alarm*-score een lagere eindscore zal toebedeeld krijgen (de betreffende participant heeft immers zijn woordenschatkennis overschat). In de discussie over de correctieformules wordt nauwelijks aandacht besteed aan hoge *Loos alarm*-scores. Men gaat ervan uit dat de *Loos alarm*-score steeds laag zal zijn want dat het aanduiden van pseudowoorden als zijnde gekend een marginaal fenomeen is. Uit ons gebruik van de Ja/Nee Toets is echter gebleken dat de *Loos alarm*-score meer dan 20% bedraagt, wat betekent dat door de bank de participanten beweren de betekenis te kennen van gemiddeld 8 van de 40 pseudowoorden. Dergelijke *Loos alarm*-scores compromitteren de betrouwbaarheid van de toetsafname en brengen bovendien de validiteit van het toetsformaat in gevaar.



Figuur 1: De stimulus-respons matrix van de Ja/Nee Toets.

2.2. De Ja/Nee toets als plaatsingstoets

Aan de Universit  Libre de Bruxelles nemen jaarlijks ongeveer 500 studenten deel aan de plaatsingstoets Nederlands. Op basis van hun score worden ze in de meest geschikte groep geplaatst bij de start van de cursus Nederlands. Deze plaatsingstoets bestond tot vrij recent uit een grammatica meerkeuzetoets die, na jaren van toetscalibratie, een betrouwbaarheid opleverde van .90. Omdat de docenten klaagden over een schrijnend gebrek aan woordenschatkennis bij de nieuwe instroom van studenten werd besloten een woordenschattoets toe te voegen aan

de plaatsingsprocedure. Er werd geopteerd voor de Ja/Nee Toets wegens de gebruiksvriendelijkheid van deze toets: het opstellen, afnemen en verbeteren van de toets neemt weinig tijd in beslag.

De gegevens die we vergaarden bij een eerste toetsgebruik werden echter gekenmerkt door een zeer hoog aantal *Loos alarm*-antwoorden (ongeveer 20%). Dit betekende dat de ruwe toetsscores sterk gereduceerd werden door het gebruik van de correctieformules. Een vergelijking van verschillende correctieformules toonde aan dat bij dergelijke hoge *Loos alarm*-scores de formules uiteenlopende resultaten opleverden. In Eyckmans (2004) brengen we verslag uit van een reeks experimenten. Daaruit komen de volgende resultaten naar voren:

1) In elk experiment (in totaal bijna 2000 participanten) werd een groot aantal *Loos alarm*-antwoorden aangetroffen (20% of meer) met de gekende gevolgen voor het transformeren van de ruwe score naar een gecorrigeerde score.

2) Bijna steeds troffen we een negatieve correlatie aan tussen de score op de woorden en de score op de pseudowoorden. Dit fenomeen kan enkel verklaard worden door de aanwezigheid van een *response bias* in de data: bij twijfel vertonen participanten de voorkeur voor één bepaald antwoordtype (*Ja* of *Nee*). Deze vertekening van het antwoordgedrag is onafhankelijk van de woordenschatkennis van de participanten.

3) Pogingen om het toetsformaat te verbeteren (i.c. het antwoordgedrag van de participanten valider te maken) aan de hand van de opgave, het computerdesign of de toetsinhoud hebben niet geleid tot een reductie of eliminatie van de *response bias*.

4) De validering aan de hand van een vertaaltaak van dezelfde woorden als die in de Ja/Nee Toets leverde slechts zeer matige correlaties op tussen de score op de Ja/Nee Toets en de score op de vertaaltoets, wat niet verbazend is gezien de aanwezigheid van een *response bias* in de data.

De conclusie van het onderzoek naar de Ja/Nee Toets is dat de *response bias* een gevolg is van een complexe interactie tussen de Ja/Nee taak op zich, het profiel van de leerder en de specifieke toetsingscontext.

2.3. De Ja/Nee taak

Het herkennen van gekende woorden en dit duidelijk maken aan de hand van een simpel *Ja* of *Nee* antwoord lijkt een eenvoudige taak, maar eigenlijk is deze opdracht zeer ambigu. Woordkennis is immers geen alles-of-niets fenomeen. Taalleerders baseren hun Ja/Nee beslissingen vaak op partiële kennis (ze herkennen de morfologie of een deel van de morfologie) of passen creatieve woordvormingstrategieën

toe, wat resulteert in het aanvaarden van teveel woord én pseudo-woord items. De Ja/Nee taak wordt vaak vergeleken met een Juist/Fout toets maar deze vergelijking gaat niet op. Bij een Juist/Fout toets is het immers mogelijk om het antwoord (en dus de kennis van de participant) te verifiëren. Wanneer iemand gevraagd wordt aan te duiden met *Juist* of *Fout* of *kaas* het Nederlandse woord is voor *fromage* en die persoon antwoordt met *Fout*, dan weten we dat de participant het woord *kaas* niet kent of althans de vertaling niet kan geven. Wanneer we hem echter vragen met *Ja* of *Nee* aan te duiden of hij de betekenis kent van het Nederlandse woord *kaas* en hij antwoordt *Ja*, dan is er geen enkele manier om na te gaan of hij die woordkennis inderdaad bezit. We kunnen hem enkel geloven. Wat dat betreft is de Ja/Nee Toets erg afhankelijk van zelfinschatting. Vooral bij twijfel levert dit ernstige problemen op omdat het beslissingscriterium in deze toets erg belangrijk is. De participant moet zich bekennen tot één van beide antwoorden – ook in geval van onzekerheid – en deze dichotomie lokt een vertekend antwoordgedrag uit. De ene leerder heeft immers meer zelfvertrouwen dan de andere, of is eerder geneigd zichzelf te overschatten dan een andere. De zelfinschatting die deel uitmaakt van de Ja/Nee taak maakt het formaat zeer gevoelig voor interactie met meta-cognitieve en socio-culturele factoren. Als gevolg hiervan zijn de Ja/Nee beslissingen geen betrouwbare weergave van de woordenschatkennis van de taalleerders. We kunnen stellen dat de meting van de woordenschatkennis in dit toetsformaat gecompromitteerd wordt door de methode, en dit is ontoelaatbaar. Daarom stellen we een nieuwe toets voor die zich kenmerkt door een andere taak.

3. Een nieuwe toets: The Recognition Based Vocabulary Test

De nieuwe toets bestaat uit itemparen van woorden en pseudoworden, waarbij de taalleerder de opdracht heeft te bepalen welk het bestaande woord is van de twee aangeboden items. Op deze manier behoudt het nieuwe formaat de algemene kenmerken van de Ja/Nee Toets (gebruik van woorden en pseudoworden, selectie uit frequentiebanden, gebruiksvriendelijk, etc.) maar wordt de “detectie-taak” van de Ja/Nee Toets in de *Recognition Based Vocabulary Test* vervangen door een “discriminatie-taak”. Deze belangrijke aanpassing van het formaat heeft verschillende voordelen:

1) In de dataverwerking kunnen geen *Loos alarm*-scores meer voorkomen. Daardoor is het mogelijk om de score van de toets op een eenvoudige en rechtlijnige manier te berekenen. Er hoeven geen complexe correctieformules gebruikt te worden om een *Treffer*-score te verlagen aan de hand van een *Loos alarm*-score. De antwoorden

zijn juist of fout en de formule die wordt gebruikt om de ruwe score te corrigeren voor gokgedrag is een klassieke *correction for guessing*-formule.

2) Er kan geen *response bias* optreden in deze toets omdat het de taal-leerder onmogelijk wordt gemaakt een voorkeur uit te drukken voor een bepaald antwoordtype. Ofwel kent de participant het correcte antwoord ofwel twijfelt hij en dan kan hij enkel gokken. Deze gokkans is voor iedereen even groot (50% in dit geval) en kan statistisch gecorrigeerd worden.

Alvorens een grootschalig experiment op te zetten, besloten we een klein pilootonderzoek uit te voeren. Hieruit bleek dat de participanten de RBVT makkelijker vonden dan de Ja/Nee Toets en dat de RBVT bovendien een grotere *face validity* bleek te hebben. De participanten gaven namelijk te kennen dat ze deze toets aantrekkelijker vonden en dat ze dachten dat dit toetsformaat hun woordenschatkennis beter zou weerspiegelen.

Er werd ook besloten een afgeleid formaat aan het experiment toe te voegen: een RBVT die bestaat uit minimale paren (*minimal pairs*), d.w.z. dat het pseudowoord is afgeleid van het woord waarmee het gecombineerd wordt. Dit afgeleid toetsformaat gaven we voorlopig de naam *Recognition Based Vocabulary Test 2*. Voorbeelden van RBVT 1 items zijn *uitkeren-overdrielen* of *straak-bioloog* (willekeurig aan elkaar gepaard) terwijl *overdrijven-overdrielen* en *biolaar-bioloog* voorbeelditems zijn van de RBVT 2.

4. Experiment

4.1. Doel

Het doel van het experiment was de twee nieuwe toetsen (RBVT 1 en RBVT 2) te evalueren en te vergelijken met de Ja/Nee Toets. Bovendien wilden we alledrie de toetsformaten valideren aan de hand van een Vertaaltoets. Zoals voorheen beschouwden we de Vertaaltoets als de meest rechtlijnige manier om de waarachtigheid van de antwoorden in de andere toetsen te verifiëren. We definiëren receptieve woordkennis in dit experiment als de vaardigheid om voor de woorden in de doeltaal één mogelijke vertaling te kunnen geven in de moedertaal. Er staat één vraag centraal: zullen de nieuwe toetsen een betere weerspiegeling zijn van de receptieve woordenschatkennis van de participanten dan de Ja/Nee Toets? Met andere woorden: welke van de formaten zal het best correleren met de Vertaaltoets?

4.2. Materiaal

Aangezien er drie toetsen werden geëvalueerd, werden er drie taalstalen getrokken (X, Y en Z). Het taalstaal werd willekeurig geselecteerd uit *Woorden in Context I en II* (Dijlts et al. 1995, 1997), een handboek waarin de basiswoordenschat van het Nederlands wordt aangebracht. We beperkten de selectie tot zelfstandige naamwoorden en werkwoorden (in verhouding tot hun voorkomen in het corpus) omdat deze makkelijker uit het geheugen kunnen worden opgeroepen wanneer men ze geïsoleerd voorgelegd krijgt (in tegenstelling tot voorzetsels of voegwoorden bijvoorbeeld). Om dezelfde redenen veronderstelden we dat ze makkelijker te vertalen waren.

Het aanmaken van de pseudowoorden gebeurde door één tot vier opeenvolgende letters te veranderen zonder de fonologische regels van het Nederlands te doorbreken. Vocalen worden niet omgezet in consonanten en vice versa.

4.3. Participanten

De participanten waren Franstalige universiteitsstudenten Sciences Economiques en Solvay Business School van de Université Libre de Bruxelles. Het waren tweede kandidatuurstudenten die de cursussen Nederlands "premier niveau" volgden. In totaal namen 177 studenten deel aan het experiment.

4.4. Methode

Het experimenteel design was opgezet rond drie formaten (Ja/Nee Toets, RBVT 1, RBVT 2) en drie materialen (X, Y, Z). Elk toetsformaat werd aangemaakt in de drie materialen. Elke participant maakte een Ja/Nee Toets, een RBVT 1 en een RBVT 2, telkens in een ander materiaal. Onmiddellijk nadien werd een Vertaaltoets ingevuld die bestond uit 20 woorden uit elk materiaal (opnieuw werd dezelfde verhouding zelfstandige naamwoorden en werkwoorden gerespecteerd).

Alle toetsen werden afgelegd op computer en de items verschenen steeds één voor één op het scherm waarbij het onmogelijk was de computertoepassing te verlaten als niet alle items beantwoord waren. Er werd geen tijdslimiet opgelegd. De opgave bij de Ja/Nee Toets luidde:

Les items qui vont apparaître à l'écran sont

-soit des mots existants en néerlandais

-soit des mots qui n'existent pas en néerlandais

Cliquez sur le bouton JA si vous pensez que le mot existe en néerlandais. Cliquez sur le bouton NEE si vous pensez que le mot n'existe pas en néerlandais.

De opgave bij de beide RBVT formaten was: *Parmi les paires d'items qui vont apparaître à l'écran, un et un seul mot existe en néerlandais. Cliquez sur le mot.*

4.5. Resultaten

Eerst werd de gelijkwaardigheid van de drie materialen geverifieerd. Hieruit bleek dat het Z materiaal uit de Ja/Nee Toets iets makkelijker was dan het X en Y materiaal, maar de verklaarde variantie was zeer klein en dit gold voor zowel de ruwe score als de gecorrigeerde scores (cfg: $F=7.016$, $df=2,174$, $p=.001$, $\eta^2=.075$; Isdt: $F=4.526$, $df=2,174$, $p=.012$, $\eta^2=.049$; Hcfb: $F=4.673$, $df=2,174$, $p=.011$, $\eta^2=.051$; ruwe score: $F=6.003$, $df=2,174$, $p=.003$, $\eta^2=.065$).

De ruwe scores op de Ja/Nee Toets (Tabel 1) waren zeer hoog en werden sterk gereduceerd door de correctieformules¹. Dit wees op een hoog aantal *Loos alarm*-antwoorden in de data.

De hoge *Loos alarm*-score werd bevestigd (zie Tabel 2): 26.92 % van de pseudowoord items werden aangekruist als zijnde bestaande woorden, dit betekent dat de participanten gemiddeld één op vier pseudowoorden beweerden te kennen. Opnieuw moet de validiteit van het toetsformaat in vraag worden gesteld. De participanten doen niet wat van hen verwacht wordt of interpreteren de taak op verschillende manieren. De betrouwbaarheden van de Ja/Nee Toets waren laag (zie Tabel 3) en de betrouwbaarheden van de pseudowoorden waren hoger dan die van de woorden. Dit betekent dat de pseudowoorden wel degelijk een systematiek vertonen binnen het toetsformaat, waarvan geen melding wordt gemaakt in de literatuur. De correlatie tussen de score op de woorden en de score op de pseudowoorden was niet significant voor de drie materialen.

¹ De correctieformules die hier worden gebruikt zijn of gebaseerd op het alles-of-niets model (cfg: correction for guessing) of op Signaal detectie modellen (Isdt en Hcfb), zie Beeckmans et al. 2001.

Ja/Nee Toets	Formule	Gem. /60	SD	%
Materiaal X (N=57)	Raw	43.90	4.74	73.17
	cfg	37.36	6.62	62.27
	Isdt	27.56	7.36	45.93
	Hcfb	23.18	6.32	38.63
Materiaal Y (N=60)	Raw	43.02	4.85	71.70
	cfg	36.02	7.59	60.03
	Isdt	27.08	8.26	45.13
	Hcfb	22.77	7.00	37.95
Materiaal Z (N=60)	Raw	46.02	5.01	76.70
	cfg	40.78	7.24	67.97
	Isdt	31.13	8.42	51.88
	Hcfb	26.29	7.24	43.82
Totaal (N= 177)	Raw	44.32	5.01	73.87
	cfg	38.06	7.41	63.43
	Isdt	28.61	8.20	47.68
	Hcfb	24.09	7.01	40.15

Tabel 1: Scores Ja/Nee Toets

		Respons alternatief	
		Ja	Nee
Stimulus alternatief	Woord	Treffer 73.87%	Misser 26.14%
	Pseudowoord	Loos alarm 26.92%	Terechte Verwerping 73.08%

Tabel 2: Matrix Ja/Nee Toets

Ja/Nee Toets	alpha (100 items)	alpha Woorden	alpha Pseudowoorden	Correlatie Woorden/ Pseudowoorden
Materiaal X (N=57)	.554	.614	.694	-.199 (p=.138)
Materiaal Y (N=60)	.627	.629	.757	-.158 (p=.229)
Materiaal Z (N=60)	.656	.665	.750	-.123 (p=.350)
Totaal (N=177)				-.134 (p=.076)

Tabel 3: Betrouwbaarheid en correlatie woorden/pseudowoorden van
de Ja/Nee Toets

De ruwe scores op de RBVT 1 waren hoger dan de gemiddelde ruwe scores op de Ja/Nee Toets (zie Tabel 4). Dit kan worden verklaard door het feit dat een discriminatietask makkelijker is dan een detectietask. De toetsen bestaan immers uit hetzelfde materiaal en elk item van de RBVT bevat dubbel zoveel informatie als een item uit de Ja/Nee Toets. Bovendien kunnen de items in een RBVT-paar dienen als elkaars referentiepunt, wat de beslissing zou kunnen vergemakkelijken. De betrouwbaarheid van de drie versies van de RBVT 1 was laag. Samen met de hoge scores zou dit kunnen wijzen op een plafondeffect en een mogelijk gebrek aan discriminerende items in de toets. De resultaten van de RBVT 2 werden gekenmerkt door nog hogere scores dan de RBVT 1 (zie Tabel 4). Dit zou erop kunnen wijzen dat het voor de participanten makkelijker is om te discrimineren tussen minimale paren dan tussen itemparen die willekeurig samengesteld zijn.

RBVT 1	Formule	/60	SD	%	alpha	RBVT 2	Formule	/60	SD	%	alpha
X (N=62)	Ruw	48.18	4.02	80.3 0	.541	X (N=58)	Ruw	50.81	3.74	84.68	.554
	Corr.	36.36	8.05	60.6 0			Corr.	41.62	7.48	69.37	
Y (N=59)	Ruw	48.08	4.58	80.1 3	.670	Y (N=58)	Ruw	51.53	3.41	85.88	.648
	Corr.	36.17	9.16	60.2 8			Corr.	43.07	6.82	71.78	
Z (N=56)	Ruw	48.64	4.08	81.0 7	.553	Z (N=61)	Ruw	52.03	3.95	86.72	.546
	Corr.	37.29	8.17	62.1 5			Corr.	44.07	7.89	73.45	
Tot. (177)	Ruw	48.29	4.22	80.4 8		Tot. (177)	Ruw	51.47	3.72	85.78	
	Corr.	36.59	8.44	60.9 8			Corr.	42.94	7.44	71.57	

Tabel 4: Gemiddelde score, betrouwbaarheid en SD van de RBVT 1 en de RBVT 2. De scores werden gecorrigeerd aan de hand van de klassieke correction for guessing-formule.

De betrouwbaarheden van de RBVT 2 waren even ontoreikend als die van de RBVT 1, wat eveneens deed vermoeden dat de toets teveel niet dicriminerende items bevatte die niet bijdroegen aan de globale toetsbetrouwbaarheid. Een itemanalyse van zowel de RBVT 1 als de RBVT 2 items bevestigde deze hypothese.

De resultaten op de Vertaaltoets waren heel wat lager dan de resultaten op de andere drie toetsen (gemiddelde score: 21/60, SD 5.16). Het daadwerkelijk kunnen leveren van een L1 vertaling voor een L2 woord bleek veel moeilijker dan de woordherkenning in de andere toetsen. De betrouwbaarheden van de Vertaaltoets waren goed

(alpha .804), wat niet verwonderlijk is aangezien het hier om een productieve taak ging. In Tabel 5 worden de correlaties getoond tussen de scores op de Ja/Nee Toets en de beide RBVT's enerzijds en de scores op de Vertaaltoets anderzijds. Een belangrijke vraag in dit experiment was immers welke van de toetsformaten het best zou correleren met de woordkennis zoals ze wordt gemeten in een vertaaltaak. De correlatiewaarden waren matig. Dit is niet verbazingwekkend gezien het gebrek aan discriminerende items in de toetsen. De lage betrouwbaarheden die dit opleverde voor de toetsen betekenen ook dat de correlatiewaarden zullen onderdrukt worden.

	Ja/Nee Formule				RBVT 1	RBVT 2
	Ruw	cfg	Isdt	Hcfb		
Vertaaltoets	.384	.538	.594	.589	.620	.689

Tabel 5: Correlatie tussen de Ja/Nee, RBVT 1 en RBVT 2 scores en de Vertaaltoets

De correlaties tussen de RBVT-scores en de Vertaaltoets waren hoger dan die tussen de Ja/Nee Toets en de Vertaaltoets, maar het verschil in deze correlatiewaarden was niet significant.

5. Conclusie

Op basis van deze resultaten kan niet worden aangevoerd dat het RBVT-formaat een grotere betrouwbaarheid en validiteit vertoont dan de Ja/Nee Toets. Wel mag worden gesteld dat de matige correlatie tussen het nieuwe formaat en de Vertaaltoets waarschijnlijk te wijten is aan een gebrek aan discriminerende items in de toets (plafondeffect) en niet aan een structureel probleem met het toetsformaat. De Ja/Nee Toets in dit experiment wordt opnieuw gekenmerkt door weinig valide antwoordgedrag. De nieuwe RBVT-toetsen lijken veelbelovender te zijn als maat voor receptieve woordenschatkennis en verder onderzoek zal moeten uitwijzen of de RBVT met willekeurig samengestelde itemparen dan wel de RBVT met minimale itemparen hiervoor het meest geschikt is. Dankzij de discriminatietaak en het feit dat het daardoor onmogelijk wordt voor de participant een *response bias* aan de dag te leggen ten overstaan van de toetsitems, zullen voorzichtige participanten in het RBVT-formaat dezelfde toetsscores behalen als te zelfzekere participanten die de neiging hebben hun woordenschatkennis te overschatten. De toetsresultaten zijn ook veel eenvoudiger te verwerken dan die van de Ja/Nee Toets, wat de RBVT gebruiksvriendelijker maakt. Verder onderzoek zou erop gericht moeten zijn de

RBVT's moeilijker te maken door bijvoorbeeld het materiaal te selecteren uit lagere frequentiebanden en de itemparen samen te stellen op basis van een moeilijkheidsindex. Het calibreren van toetsen lijkt wat dat betreft helemaal terug van weggeweest. Uiteraard merken we ook op dat de RBVT's slechts een soort elementaire vorm van L2 woordkennis meten en het is duidelijk dat de productieve taak van de Vertaaltoets een andere vaardigheid vergt. Daarom moet er voorzichtig omgesprongen worden met de extrapolatie van toetsscores naar de totale L2 woordenschatomvang van een taalleerder.

Bibliografie

- Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M. & Van de Velde, H. (2001) Examining the Yes/No vocabulary test: some methodological issues in theory and practice. *Language Testing*, 18, 235-274.
- Coady, J., Magoto, J., Hubbard, P., Graney, J. & Mokhtari, K. (1993) High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes and J. Coady (Eds.) *Second Language Reading and Vocabulary*. Norwood, NJ.: Ablex, 3-23.
- Dialang (2003) <http://www.dialang.org>
- Dieltjens, L., Vanparijs, J., Baten, L., Claes, M.-T., Alkema, P. & Lodewick, J. (1995) *Woorden in Context Deel 2*. Brussels: De Boeck.
- Dieltjens, L., Vanparijs, J., Baten, L., Claes, M.-T., Alkema, P. & Lodewick, J. (1997) *Woorden in Context Deel 1*. Brussels: De Boeck.
- Eyckmans, J. (2000) De Ja/Nee woordenschattoets: klaar voor gebruik in de klas? *Toegepaste Taalwetenschap in Artikelen*, 2, 117-128.
- Eyckmans, J. (2004) *Measuring Receptive Vocabulary Size: Reliability and Validity of the Yes/No Vocabulary Test for French-speaking learners of Dutch*. Utrecht: LOT.
- Hirsh, D. & Nation, I.S.P. (1992) What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689-696.
- Huibregtse, I., Admiraal, W & Meara, P. (2001) Scores on a yes/no vocabulary test: correction for guessing and response style. *Language testing* 18, 227-245.
- Laufer, B. (1992) How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan, 126-132.
- Meara, P. (1990) Some notes on the Eurocentres vocabulary tests. In J. Tommola (ed.) *Foreign Language Comprehension and Production*. Turku: AFinLa, 103-113.
- Meara, P. (1992) *EFL Vocabulary Tests*. Swansea: Centre for Applied Language Studies, University of Wales.
- Meara, P. (1996) The dimensions of Lexical Competence. In G. Brown, K. Malmkjaer and J. Williams (eds.) *Performance and Competence in Se-*

- cond Language Acquisition*. Cambridge: Cambridge University Press, 35-53.
- Meara, P & Buxton, B. (1987) An alternative to multiple choice vocabulary tests. *Language Testing*, 4, 142-151.
- Nation, I.S.P. (1983) Testing and teaching vocabulary. *Guidelines* 5, 12-25.
- Nation, I.S.P. (1990) *Teaching and Learning Vocabulary*. New York : Heinle and Heinle.
- Nation, I.S.P. (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Schmitt, N. (2000) *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.